

**Technical Report
1118**

Performance Assessment of Intelligence, Surveillance, and Reconnaissance (ISR) Enterprise Workflows

**M.B. Hurley
P.B. Jones**

31 January 2007

Lincoln Laboratory
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
LEXINGTON, MASSACHUSETTS



Prepared for the Department of the Navy under Air Force Contract FA8721-05-C-0002.

Approved for public release; distribution is unlimited.

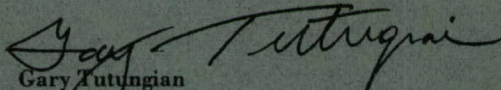
This report is based on studies performed at Lincoln Laboratory, a center for research operated by Massachusetts Institute of Technology. This work was sponsored by the Department of the Navy, Office of Naval Research, under Air Force Contract FA8721-05-C-0002.

This report may be reproduced to satisfy needs of U.S. Government agencies.

The ESC Public Affairs Office has reviewed this report, and it is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nationals.

This technical report has been reviewed and is approved for publication.

FOR THE COMMANDER


Gary Tutungian
Administrative Contracting Officer
Plans and Programs Directorate
Contracted Support Management

Non-Lincoln Recipients

PLEASE DO NOT RETURN

Permission has been given to destroy this document when it is no longer needed.

**Massachusetts Institute of Technology
Lincoln Laboratory**

**Performance Assessment of Intelligence, Surveillance, and
Reconnaissance (ISR) Enterprise Workflows**

*M.B. Hurley
P.B. Jones
Group 104*

Technical Report 1118

31 January 2007

Approved for public release; distribution is unlimited.

Lexington

Massachusetts

ABSTRACT

One of the challenges of developing complex systems is the question of how to measure system performance. Good engineering practice is to define a single compound measure of system performance, called the figure of merit (FOM), and to select design options by their effects upon the figure of merit. The Joint Services of the Department of Defense (DoD) are in the process of transitioning legacy computing applications for intelligence, surveillance, and reconnaissance (ISR) tasks to standards-based information enterprises. This process is occurring with very little theoretical or practical understanding of the appropriate FOMs needed to assess the performance of these large information systems. The danger of developing systems without clearly stated and understood FOMs is that important but difficult to measure system characteristics are overlooked in favor of less important but easier to measure system characteristics. As a result, operational users, system developers, and program managers who rely on accurate performance information are left with considerable uncertainty regarding how to best use, improve, and deploy these systems.

This report presents an end-to-end assessment framework for C4ISR enterprises that identifies potential FOMs and thereby eliminates a significant fraction of this uncertainty. The framework delineates how measurements from laboratory tests, operational experiments, and field deployments can be analyzed to produce accurate, quantitative descriptions of system performance. These FOMs not only inform the development and acquisition processes but also allow military users to ensure that they are optimizing the information enterprise to achieve improved operational capability. In fact, active participation by the military user community in defining appropriate FOMs is essential to providing users with systems that meet their operational needs. An analysis of a simple ISR enterprise is conducted to show how FOMs might be calculated and used to assess the relative merit of different communications architectures for the construction of a common operational picture (COP) among a collection of distributed sensors.

TABLE OF CONTENTS

Abstract	iii
List of Illustrations	vii
List of Tables	vii
1. INTRODUCTION	1
1.1 Motivation	1
1.2 Method	2
1.3 Report Organization	2
2. EVALUATION OF SYSTEMS	3
3. OTHER ASSESSMENT MODELS	5
3.1 MIT Lincoln Laboratory: Silent Hammer Limited-Objective Experiment Analysis	5
3.2 Office of Force Transformation: Network-Centric Operations Conceptual Framework	6
3.3 Perry, Signori, Boon: Shared Awareness	7
3.4 Moffett: Complexity Theory for Network-Centric Warfare (NCW)	8
3.5 Kadambe and Daniell: Assessment of Distributed Networks	8
3.6 Endsley: Situation Awareness	8
3.7 Mahoney, Laskey, Wright, Ng: Situation Assessment	8
3.8 Various Information Quality Papers	9
3.9 Literature Search Summary	11
4. A SIMPLE DECISION MODEL	13
4.1 The First-Order Decision Model	13
4.2 The Second-Order Decision Model	16
4.3 The Third-Order Decision Model	21
5. A DECISION MODEL FOR ISR ENTERPRISES: A SIMPLE DEMONSTRATION	27
6. CONCLUSIONS	35

TABLE OF CONTENTS (CONTINUED)

APPENDIX A: METRICS FOR OTHER MODEL ATTRIBUTES	37
A.1 Accuracy	37
A.2 Consistency	37
A.3 Security	38
A.4 Latency and Timeliness	38
A.5 Completeness	39
A.6 Conciseness	39
A.7 Reliability	39
A.8 Accessibility	39
A.9 Availability	40
A.10 Objectivity	40
A.11 Relevance	40
A.12 Usability	40
A.13 Understandability	41
A.14 Amount of Data	41
A.15 Believability	41
A.16 Navigation	41
A.17 Reputation	42
A.18 Usefulness	42
A.19 Efficiency	42
A.20 Value-Added	42
References	43

LIST OF ILLUSTRATIONS

Figure No.		Page No.
1	The first-order decision model.	14
2	The second-order decision model.	17
3	The third-order decision model.	21
4	Demonstration of sensing capability for sensor B.	30
5	Average Shannon entropy, KL distance, and Bayesian agreement with truth for each of the five sensors when using the “unlimited communication” architecture.	31
6	Average Shannon entropy, KL distance, and Bayesian agreement with truth for each of the five sensors when using the “no communication” architecture.	31
7	Average Shannon entropy, KL distance, and Bayesian agreement with truth for each of the five sensors when using the “blind push” architecture.	32
8	Average Shannon entropy, KL distance, and Bayesian agreement with truth for each of the five sensors when using the “blind pull” architecture.	32
9	Average Shannon entropy, KL distance, and Bayesian agreement with truth for each of the five sensors when using the “informed pull” architecture.	33

LIST OF TABLES

Table No.		Page No.
1	System Characteristics for Network Centric Operations	6
2	A Compilation of Information Quality Attributes	10
3	Sensor Detection and Discrimination Strengths	29
4	Sensor Observation Extent	29

1. INTRODUCTION

MIT Lincoln Laboratory, in cooperation with the Distributed Common Ground System–Navy Program Office (DCGS-N), has conducted research to develop a rigorous methodology for characterizing the performance of service-oriented intelligence, surveillance, and reconnaissance (ISR) enterprises. The Distributed Common Ground System (DCGS) is being developed in parallel by the four Joint services to provide a common environment for the management and exploitation of ISR data throughout the Department of Defense (DoD). The DCGS system is designed as a service-oriented architecture (SOA) and its performance will be difficult to assess using traditional performance characterization techniques. DCGS provides a modular software environment in which discrete application “services” can be used independently or composed into more sophisticated “workflows.” Characterizing the performance of individual services ignores the emergent characteristics of workflows. On the other hand, exhaustively assessing workflows can be impractical due to the number of service combinations possible in a large enterprise. These problems are exacerbated by the fact that the DCGS system will support rapid upgrade cycles for the constituent services. An enterprise assessment framework must not only characterize service-level performance but also potential workflow interactions between services. Furthermore, the assessment framework should support rapid, *in situ* characterization of new or improved services and workflows to allow rapid development of enhanced capabilities.

This effort focused on identifying a theoretical basis upon which ISR processing and exploitation services can be assessed individually and in the context of larger workflows. The program involved three interrelated tasks: a multidisciplinary study of enterprise systems assessment, development of an assessment methodology specific to ISR enterprises, and, finally, the application of that methodology to a simple simulation of an ISR enterprise with relevance to the DCGS-N program. The application developed for this study is a simulation of multiple, distributed, heterogeneous sensors attempting to develop a common operational picture (COP). The analysis examines the impact of different communications architectures on the commonality and accuracy of the COP as the sensors observed the environment.

1.1 MOTIVATION

Prior techniques for evaluating enterprise information systems fall mostly into the category of heuristic methods. Expert knowledge, rules of thumb, and functional decomposition are used to first identify figures of merit (FOMs), then derive figures of effectiveness and finally select figures of performance for the problem space. While these approaches have performed well in the past, they rely heavily on the experience and expertise of subject matter experts to be effective. In addition, it is often difficult to adapt the methodology developed for one system to another system. A more general analytic foundation for assessment can significantly improve the ability to evaluate novel enterprises and provide for an easier transition of experience from one analysis to another.

1.2 METHOD

This project selected Bayesian probability theory and information theory as general constructs for enterprise performance assessment. The central postulate of the analysis is that ISR enterprises, like DCGS-N, can be modeled as a distributed decision system made up of a set of interconnected, elementary decision systems. The activity of distributed decision systems is dominated by the information flow between the multiple, loosely coupled components that form the system.

This view of ISR enterprises was selected because the fundamental purpose of ISR processes is to collect and concentrate mission-relevant data to support operational decision-makers. Probability theory and information theory were adopted as the mathematical foundations because the theories provide objective measures of information as it flows through communications networks. The distributed decision system model is constructed with an analytical foundation that guides the selection of FOMs for ISR missions and that are, ultimately, still motivated by expert knowledge and the needs of the warfighter.

This effort was supported by a survey of systems modeling research reports to identify alternate analytic techniques for enterprise assessment, besides probability theory, information theory, and statistical decision theory. An incomplete list of research areas that were surveyed includes formal methods, pi-calculus, domain theory, I/O automata, automated theorem proving, and the unified situation modeling language. In some cases, the connection between these theories and distributed services architectures was found to be weak or obscure. A few theories looked promising but were not at a sufficient level of maturity to provide figures of merit. This was certainly the case for pi-calculus, which has been developed to model decentralized systems and complex business processes. The current focus of research in pi-calculus is to develop modeling techniques for these systems and processes and has not yet begun to focus on evaluation techniques. This may be a research area to continue to monitor as it transitions into systems evaluation.

1.3 REPORT ORGANIZATION

The remainder of the report will be organized as follows. Section 2 discusses the principles of system evaluation. Section 3 presents a survey of suggested evaluation techniques. In Section 4, a model of distributed decision systems (such as the ISR Multi-INT enterprise) will be developed. In Section 5, the application of probabilistic and information theoretic measures to the model will be demonstrated, first in theory and then through the use of a “toy” ISR enterprise. Finally, Section 6 contains conclusions and suggestions for future work.

2. EVALUATION OF SYSTEMS

The question that motivated this analysis was whether a single FOM could be developed for ISR enterprises based upon SOAs, like DCGS-N. For our purposes, a FOM is a measure that captures the overall value of a system to the end-user. FOMs can be decomposed into measures of effectiveness (MOEs), which can in turn be decomposed into measures of performance (MOPs). Examples of well-known FOMs are the radar equation and the sonar equation. The power of these FOMs is that they allow the system designer to quantify overall system trade-offs under variations of disparate system parameters. The central question of the research was whether a FOM could be identified for SOAs, thereby providing a tool to evaluate the relative performance of architectural variations within the ISR enterprise.

One objective measure that immediately came to mind as a potential FOM was the total life-cycle cost of a system. The comparison of costs between systems would provide an objective means of assessing the value of one system over another. Obviously, the total life-cycle cost of a system is related to the goals that an evaluator expects a system to achieve. The practical estimation of total life-cycle cost is often difficult because costs need to be assigned to all aspects of the system and the environment in which the systems operate. Although difficult, the financial world and government acquisitions programs attempt to estimate total life-cycle costs on a continual basis.

If FOMs are associated with total life-cycle cost, then MOEs can be naturally associated with the component costs in the overall system that together provide an estimate of total life-cycle cost. MOPs are naturally associated with physical measures such as time, position, velocity, energy, power, work, and temperature. Many analysts define MOEs as differences and ratios of MOPs and adopt these as FOMs. In fact, the radar and sonar FOMs are of this class. If cost is really the dominant measure for FOMs, then these analysts make an implicit assumption that differences and ratios of MOPs are directly proportional to differences and ratios of costs. Of the three different forms of evaluation metrics, the comparison of physical measures for different systems is the most common because physical measurements are the easiest to collect and do not involve cost estimates. However, these comparisons may not be truly indicative of the true value of different systems because costs may not be directly proportional to the MOPs.

3. OTHER ASSESSMENT MODELS

To avoid the need to create a new model for service-oriented architectures, the technical literature was searched to identify prior models that might be used in the evaluation of the ISR enterprise. The number of prior studies that were directly related to either ISR systems or service-oriented architectures was limited to a handful of articles and books. A few other applicable studies were found for related topics such as situation assessment. No study discussed the development of a mathematical foundation for the evaluation of systems. In all cases examined, heuristic arguments or expert opinions were used to select measures of performance and measures of effectiveness. No study was located that attempted to develop a single figure of merit.

3.1 MIT LINCOLN LABORATORY: SILENT HAMMER LIMITED-OBJECTIVE EXPERIMENT ANALYSIS

A paper that describes the analysis of the Silent Hammer Limited Objective Experiment [1] was examined as a starting point for the literature search. Colleagues of the authors of this analysis were involved in Silent Hammer and available for technical discussions. The analysis was directed at answering several overarching questions related to the operational capabilities of submersible, ship, guided, nuclear (SSGN) submarines for conducting future missions. The three pillars of the Navy Sea Power 21 vision (Sea Strike, Sea Shield, Sea Basing) formed the basis of the decomposition of the questions. The Navy's Mission Capability Packages (MCPs) were selected from the three pillars with regard to their relevance to the Silent Hammer experiment and to define the operational capabilities that the analysis would examine. The Navy Uniform Joint Task List was analyzed and the tasks that were relevant to Silent Hammer were selected for each MCP. Mission scenarios were constructed that would exercise those tasks during the experiment. Operational capabilities were defined for each mission in correspondence to the mission capability packages. The operational capabilities fell into one of four different analysis categories:

1. Information production
2. Information distribution and dissemination
3. Processing and decision
4. Action/execution

MOEs were then defined for each operational capability, with the mission scenarios providing guidance as to the importance of the different measures. These measures fell into one of four different MOE categories:

1. Timeliness
2. Quality
3. Completeness
4. Risk reduction

The analysis plan phrased each measure of effectiveness as a question relating to an aspect of an operational capability. MOPs were then identified for each question. In all cases, the MOPs were derived from actual physical parameters.

Although the analysis was able to definitively answer all the questions in Silent Hammer and provide analytic justification for the answers via the measures of performance, no mathematical foundation was used to guide the quantitative analysis of each question.

3.2 OFFICE OF FORCE TRANSFORMATION: NETWORK-CENTRIC OPERATIONS CONCEPTUAL FRAMEWORK

The Office of Force Transformation produced the next document that was examined for information on the creation of a detailed conceptual framework for network-centric operations [2]. The framework spans three domains: the physical domain (ground truth), the information domain, and the cognitive domain. The top level of this model enumerates several characteristics of the conceptual framework. These characteristics are summarized in Table 1. The next level defines attributes and metrics that are mapped to the system characteristics enumerated at the higher level.

Table 1
System Characteristics for Network Centric-Operations

Quality of:	Degree of:	Agility
Networking	Information "shareability"	C2 agility
Organic information	Shared information	Force agility
Individual information	Shared sense-making	
Individual sense-making	Decision synchronization	
Interactions	Actions/entities synchronized	
	Effectiveness	

The concepts are linked in the objective framework with the understanding that improvements in one characteristic will most likely lead to improvements in the other characteristics.

The conceptual framework defines objective and fitness measures for each top-level concept. The goal is for these measures to be analytic and objective wherever possible, but recognizes that many measures will have to be qualitative. For example, the objective measures for the quality of organic information are

- Correctness
- Consistency
- Currency
- Precision

The framework defines “fitness-for-use” measures to be

- Completeness
- Accuracy
- Relevance
- Timeliness

The framework provides specific measures and ratios for each category. The developers of the conceptual framework view it as a work in progress that will be modified as more knowledge is gained about network-centric operations. Although the conceptual framework is well known throughout the C4ISR community and is very detailed, it is not built off a fundamental mathematical framework.

3.3 PERRY, SIGNORI, BOON: SHARED AWARENESS

In the study conducted by Perry et al. [3] at RAND, the Network-Centric Operations Conceptual Framework was adopted as the foundation for a study on a methodology for measuring the quality of information and its impact on shared awareness. The study had the explicit goal of developing a mathematical framework to facilitate the development of MOPs and other metrics for assessing situational awareness. The model that was examined was a specific class of C4ISR architectures in which an array of sensors collects measurements on the environment and reports them to a fusion center that constructs a common relevant operational picture (CROP). Relevant aspects of the CROP are then distributed to members of the network.

The study limits further analysis to a subset of measures of the quality of organic information: correctness, completeness, and currency. The authors use estimation theory to derive quality measures based upon estimation theory. Despite the mathematical foundation and the ability to collect quality measures together for unified measures that could capture overall quality and stand in for a FOM, the rigor of the mathematical foundation was found to be insufficient for systematically assessing ISR enterprises and SOAs.

3.4 MOFFETT: COMPLEXITY THEORY FOR NETWORK-CENTRIC WARFARE (NCW)

The review of the RAND study lead to the discovery of a recently published book by James Moffett on complexity theory and network-centric warfare [4]. The introduction of the book points out that complexity theory is not yet a mature, well-defined theory but is instead constructed from a number of different theories, including information theory, in an attempt to characterize complex systems. The book uses Shannon information to define a measure of residual information and a measure of detection knowledge with the residual information equation and other information-theoretic equations. The selection of information theory as a mathematical foundation aligns with the views of the authors of this study, but the book does not provide a conceptual model of network-centric systems to which the mathematical foundation is to be applied.

3.5 KADAMBE AND DANIELL: ASSESSMENT OF DISTRIBUTED NETWORKS

Kadambe and Daniell [5] at Hughes Research Laboratory conducted research into theoretic foundations for the performance assessment of distributed sensor networks. They examined the applicability of Euclidean distances, Shannon information, mutual information, and symmetric Kullback-Leibler distances for assessing the performance of a distributed sensor network application. In addition to distance measures, energy costs were proposed as an important measure of system performance, specifically in the energy required to conduct measurements. The distance measures were used to assess the value of contributing measurements in the fusion process and to determine if measurements should be fused. Results focus on the performance of classifiers that reject measurements based upon an assessment of the value of information provided by the measure to the fused result. If the result would degrade the quality of the classification decision, the measurement was not fused.

This report provides for measures that the authors of the study believe are the correct ones to adopt for an analysis of service-oriented architectures, but does not provide either a model to drive the selection of the measures or a fundamental FOM.

3.6 ENDSLEY: SITUATION AWARENESS

The research papers of Mica Endsley [6, 7] are frequently referenced because of her model for describing situation assessment. These papers describe an interesting and useful model for designing and understanding human-centric situation assessment, but the model was not constructed with a mathematical foundation.

3.7 MAHONEY, LASKEY, WRIGHT, NG: SITUATION ASSESSMENT

Due to the influence of the Endsley papers on human factors and situational awareness research, a search was conducted for papers on performance metrics for situational awareness. A paper by Mahoney et al. [8] was found that presents evaluation measures for situation assessment systems. In the paper, situations are modeled as probability density functions (PDFs) over hypothesized groups, units, sites, and

activities. The situation assessment systems report PDFs for the possible compositions of groups, units, sites, and activities. Situation PDFs are scored through the comparison of the PDFs with ground truth. Scoring can be carried out at the level of individual components to the level of the global situation. Additional techniques are presented for scoring systems when ground truth is not available.

Users can specify their attributes of interest in the situation and the relative importance of these attributes. Four aspects of a situation assessment score are given as

- Fidelity of individual system elements (vector of scores)
- Overall fidelity
- Value of the element to the user
- Utility: the value of element versus the cost of the element

The elements of the fidelity vector are selected by heuristic means and the overall fidelity score is a function that combines the elemental scores into one value. Bayesian networks are used to automate the generation of scores, with a network being constructed for each user's query. Although probability theory and information theory are used to generate some of the numbers that factor into the scores, the actual functions selected to generate the scores do not conform to the axioms of probability theory.

The paper notes that PDFs vary through space and time and that the temporal variation in the PDFs can be included in the scores. The authors report on three different approaches to scoring time dependent PDFs:

1. Compare estimates and ground truth at the times of validity for generated reports,
2. Compare estimates and ground truth across of an interval of time, and
3. Compare estimates and ground truth as a series of snapshots in time.

This material is particularly relevant to our current research in that situations are represented by PDFs and the suggested measures account for the temporal nature of the PDFs. However, even though the system under evaluation uses Bayesian networks to conduct probabilistic reasoning, the measures that are adopted to score system performance are not formulated as correct probabilistic expressions. The models that are discussed are models of the system instead of models of the evaluation process.

3.8 VARIOUS INFORMATION QUALITY PAPERS

A final area of investigation was the field of information quality, which is primarily an area of study in the business research community. Multiple researchers have defined several aspects describing the information quality of data stored in information systems. The most recent and most informative overview of the measures in this research area is by Knight and Burn [9]. Their compiled list of information quality dimensions is shown in Table 2. The target for their research is automated techniques to assess the quality of information found by Internet search engines. They cite other researchers who state that information quality cannot be assessed without considering the context of its generation and intended use.

The general framework for information quality assessment requires that three entities be identified and understood: the user, the environment, and the task. The three entities are used to rank the dimensions of information quality. No detailed mathematical foundation is provided for the framework.

Table 2
A Compilation of Information Quality Attributes [9]

Measures	Definitions
Accuracy	Extent to which data are correct, reliable, and certified free of error
Consistency	Extent to which information is presented in the same format and compatible with previous data
Security	Extent to which access to information is restricted appropriately to maintain its security
Timeliness	Extent to which the information is sufficiently up-to-date for the task at hand
Completeness	Extent to which information is not missing and is of sufficient breadth and depth for the task at hand
Conciseness	Extent to which information is compactly represented without being overwhelming (i.e., brief in presentation, yet complete and to the point)
Reliability	Extent to which information is correct and reliable
Accessibility	Extent to which information is physically accessible
Availability	Extent to which information is physically accessible
Objectivity	Extent to which information is unbiased, unprejudiced, and impartial
Relevancy	Extent to which information is applicable and helpful for the task at hand
Usability	Extent to which information is clear and easily used
Understandability	Extent to which data are clear, i.e., without ambiguity and easily comprehended
Amount of data	Extent to which the quantity or volume of available data is appropriate
Believability	Extent to which information is regarded as true and credible
Navigation	Extent to which data are easily found and linked to
Reputation	Extent to which information is highly regarded in terms of source or content
Usefulness	Extent to which information is applicable and helpful for the task at hand
Efficiency	Extent to which data are able to quickly meet the information needs for the task at hand
Value-Added	Extent to which information is beneficial, provides advantages from its use

3.9 LITERATURE SEARCH SUMMARY

To date, no documents have been discovered to provide a mathematically sound foundation for the evaluation of ISR enterprises and service-oriented architectures. Given that these architectures are inherently network-centric, a number of relevant papers were found on frameworks and models for network-centric systems. Many of these papers focused specifically on the C4ISR enterprise, which is exactly what DCGS-N is. Some partially relevant papers were found in other research areas, including situation awareness and assessment, and information quality. Common measures throughout many of the papers were quality, timeliness, accuracy, and completeness. The information quality field apparently has the more imaginative researchers and has produced twenty metrics. For those papers that propose mathematical expressions for the measures, the expressions are selected based upon heuristics and many are selected by ad hoc methods. In most cases, the measures are not based upon a firm theoretical foundation and often the ad hoc methods lead to expressions that a mathematical foundation would cause one to reject.

A number of gems were found in the literature search. The first gems were models of decision systems that looked sensible, primarily the Endsley model, with a decision system existing in an environment that it senses and modifies. This theme was found in a number of other papers. Secondly, measures based upon probability theory and information theory were found in several papers and present a promising avenue of investigation. Unfortunately, these papers seldom focus on how the measures related to models of the evaluation process.

Based on the negative results of the search for a well-founded, mathematically based evaluation method for Multi-INT enterprises, a second information gathering phase occurred in which several mathematical theories were investigated for applicability to the problem of Multi-INT enterprise evaluation. Investigated theories included probability theory, information theory, statistical decision theory, formal methods, pi-calculus, domain theory, input/output (I/O) automata, automated theorem proving and the unified situation modeling language. Probability theory, information theory, and statistical decision theory were selected to form the mathematical foundation for the model of an evaluation process. This selection was partially motivated by the maturity of these theories, the inter-relationships between these theories, the selection of measures from these theories by other researchers, and the familiarity that the authors of this study already have with these theories.

4. A SIMPLE DECISION MODEL

The models of other researchers were informative, but did not have suitable structures for the construction of a model with a mathematical foundation. In order to provide an adequate mathematical foundation, a new series of simple models was constructed. The new models provide analytic and objective figures of merit, measures of effectiveness, and measures of performance. The series of models was constructed by starting from the simplest possible model (sense, decide, actuate) and then progressively adding pertinent details at each step of model construction. The mathematics supporting the models was integrated into the first model and continually expanded with each additional model as each increment increased in complexity. Although the research in this study was focused specifically on the performance assessment of the DoD's ISR enterprises, the resultant model is applicable to a larger set of decision systems.

4.1 THE FIRST-ORDER DECISION MODEL

The simplest description of a decision system is that it exists in an environment and has sensors that measure aspects of (or perceive) the environment and actuators that can change the environment. The nature of the measurements that the decision system makes may induce it to use its actuators to alter the environment. The minimum mathematical description for this model is a mapping from sensor measurements to actuator actions:

$$A = T(S), \tag{1}$$

where A is the determined action, selected via the transformation function T and the sensed data S . This simple model is shown in Figure 1. This decision system forms the fundamental unit for the more complex decision models that follow.

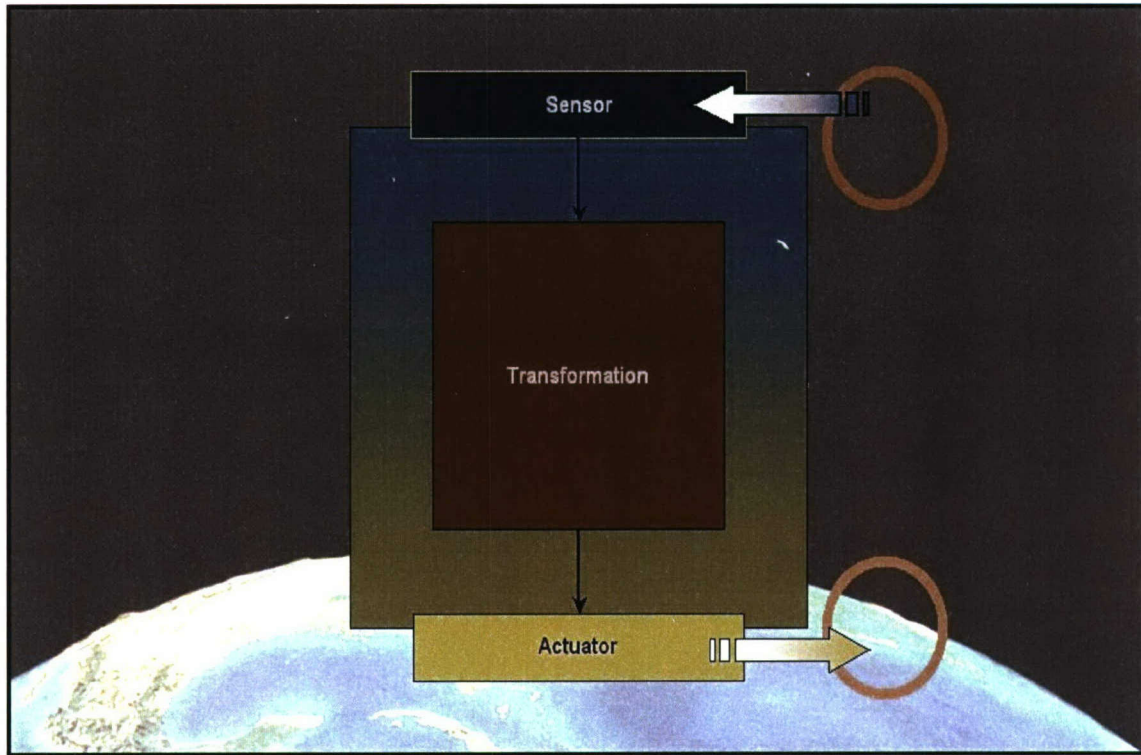


Figure 1. The first-order decision model.

The decision system and its environment can be modeled as a state machine, in which the environment exists in a specific state (the “world state”) and the decision system senses a substate of the world state and then uses its actuators to change a (possibly different, or even disjoint) substate of the world state. The environment’s state may be changing over time, depending on the physics of the environment. Also, because the system is embedded in the environment, its state should be considered to be part of the world state. The observability of the system’s internal states will constrain what measures can be collected for evaluation.

When considering the evaluation of the decision system in its environment, an additional component may be added to the model in the form of an evaluator. The evaluator’s goal is to assess the decision system’s performance in its environment. In essence, the evaluator is a second decision system with the goal of evaluating other decision systems. The full description of the evaluator is more complex than the first-order decision model will support, but the second-order model that will be described later can support it.

The addition of an evaluator to the environment facilitates a number of different model configurations. The evaluator can be embedded in the environment or the evaluator can be modeled as outside the environment, looking in. The evaluator may be capable of sensing the entire world state or may be limited to sensing only a portion of it. If the evaluator is limited to sensing only a portion of the state, this portion may or may not include the subspace that the decision system can sense and act upon. The evaluator may or may not be able to sense the subspace that consists of all or part of the internal state of the decision system. Different configurations lead to different measures that the evaluator can collect to evaluate the performance of the decision system.

One of these configurations will be examined to illustrate how an evaluator might judge the performance of a decision system. For this configuration, the evaluator has no knowledge of the internal states of the decision system, nor of how actions are selected. The evaluator can only measure the part of the world state that is external to the decision system, meaning the evaluator can only assess the performance of the decision system by measuring the overall effect of its actions on the rest of the world. For this configuration, the evaluator can be considered to have the ability to set the environment to an initial state and observe the evolution of the environment across repeated tests. An evaluator can make two types of comparisons: 1) comparisons between the evolution of the environment with the decision system and without, and 2) comparisons between different decision systems. (If the environment without a decision system is considered to contain a null decision system, the comparison can be considered to be of only one type.) Measures of performance can be derived from the direct measurements that the evaluator collects, such as the proportion of the environment that the decision system can change; the speed at which the decision system changes the environment's states; the states, if any, that the decision system drives the environment toward; and the latencies between changes in environmental states and the decision system's responses to the changes.

The repeated trials provide the mechanism for the evaluator to collect performance statistics. The statistics can be used to estimate probabilities, $P(E, t)$, for the system to achieve a given environmental state, E , over the course of a run (t is time). The principle of maximum entropy [10], written mathematically in Equation (2), can be used to estimate the probabilities from the collected measurements, x , and prior information, i .

$$P_{ME}(E | x, i) = \arg \max_P (P(E | x, i) \ln(P(E | x, i))). \quad (2)$$

Intuitively, the principle of maximum entropy chooses, from among all the possible probability density functions that match the observed data, the one that makes the fewest additional assumptions. Actual application of the maximum entropy techniques can be challenging and the techniques for generating the PDFs vary for different environments. A detailed description of maximum entropy techniques is beyond the scope of the study, but is only proposed as a reasonable method to obtain PDFs for the evaluation process.

In the decision model, evaluators can conduct additional analyses if they have a value or cost function for the environmental states in addition to the probabilities that are derived from the results of

the trials. The value function is a reflection of the goals of the evaluator and is used to generate measures of effectiveness and a figure of merit, which may or may not reflect the goals of the decision system. The expected average value of the decision system can be estimated using the value function, $C(E)$, and the derived PDF using Equation (3).

$$\langle C \rangle = \frac{\int_{t_{start}}^{t_{end}} \sum_E C(E) P(E | x, i, t) dt}{t_{end} - t_{start}}. \quad (3)$$

If the evaluator possesses exact knowledge of the true world state, that knowledge is represented by a PDF that is a delta function and the expected value is the value of the true world state.

The nature of the expected value of decision systems leads to the conclusion that no universal FOM can exist for an environment with decision systems because, in general, different evaluators will have different goals that will be reflected in the different values that they assign to the environmental states.

Additional configurations of the model at this level of complexity could be considered, such as different capabilities for the evaluator to sense or control the environment. Those configurations might be valuable in developing actual evaluation systems. The principal interest of almost all evaluators eventually turns toward understanding the internal states of the decision system itself, and so the complexity of the model will be increased to describe these kinds of evaluations.

4.2 THE SECOND-ORDER DECISION MODEL

The motivation for moving to a more complex model is to develop a model of an evaluation process for distributed service-oriented architectures. The tenet of this report is that these large-scale complex architectures either describe decision systems or components in a greater decision system and that the true value of these architectures can only be evaluated from the perspective of the complete decision system. The detail of the model thus far does not provide insight into how decision systems function. This next step adds a basic description of the internal functions of decision systems and is shown in Figure 2. As was the case for the simplest model shown in Figure 1, this model has the decision system, with sensors and actuators, interacting with an environment. This model shows the internal data and processes that form the components of a simple decision system. This second-order decision process can be seen to be composed of fundamental units that are equivalent to the single unit in the first-order model. This unit is a mathematical transformation that receives input data and produces output data.

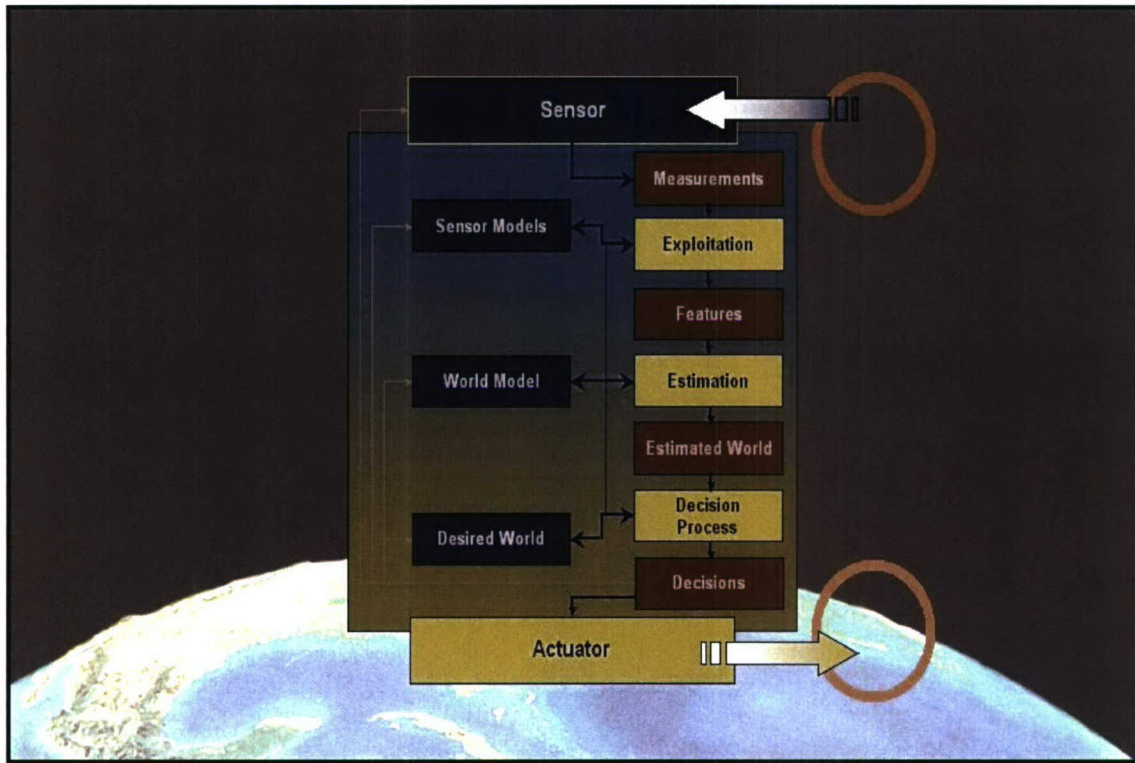


Figure 2. The second-order decision model.

The first step in the decision process is a detection step in which the sensors sense the environment and generate measurements. This transformation may rely on a sensor model to transform actual sensed data to measurements that are provided to the next step in the processing chain. For example, radars sense incident intensities but report radar cross-section or target detections.

The second step in the process is an exploitation step that converts measurements into features. Features are often used to more accurately or succinctly capture the parameters that are relevant to either the system's model of the environment or desired environmental states (goals). The figure shows that the exploitation step may be dependent on the system's sensor model and environmental model. This represents system knowledge with respect to the system's sensing process and the environment.

The third step is a process that generates estimates of the environment's state from the features. This step uses an environmental model to convert the features to estimates of the world state.

The fourth step is a decision process that uses the estimated world state to select the actions that will transition the real world state to a desired world state. Mathematically, this step is a mapping between

the estimated world state and the actions that will transition the world state to a more desired one. The system's model of desired world states represents the goals of the system. As mentioned in the section describing the first-order model, the first-order model lacked sufficient complexity to model an evaluator because it lacked a method of modeling "goals" of the system. The second-order model explicitly depicts the role of goals in a decision system, facilitating the more accurate modeling of the evaluator.

The last step is the actuator process that receives the decisions and transitions the environment to another state. This process can also change the internal states of the decision process to either improve the accuracy of the system's models or change the goals that the system is striving toward. This ability provides a means for the decision system to learn.

A number of models that were found in the technical literature search influenced the design of this model. Mica Endsley's model [6] had a significant impact on the design of this model. Whereas her model is focused on situation assessment and issues of human-computer interactions, this model is focused on the formulation of a mathematical foundation for system evaluation. The model of interacting agents that is described by Leslie Kaelbling et al. is very much like the model described here and influenced the design as well [11]. The Kaelbling team focused on how to build agents (or decision systems) to operate effectively in an environment. Their focus is on how agents can improve performance through learning rather than on the evaluation of the agents' performance. Another influence on the design was a paper by John Campion [12], which highlights two related problems in the psychology community: 1) the divide between the naturalistic decision-making camp and the classical decision-making camp and 2) the separation of control into subareas of decision-making and situation awareness. He argues for a unified model that captures the entire enterprise, not just the components.

The transformation between measurements and features could have been left out of the model at this point to slightly decrease the complexity of the model at this level. It was retained for two reasons: to highlight the fact that the primary function of many processes in a decision system is to compress or convert data to more convenient forms for other processes and eventually to provide a more faithful mapping to the C4ISR architecture in the Navy.

4.2.1 Evaluation of Second-Order Decision Models

As with the first-order model, a number of evaluation configurations can be constructed. A simple configuration will be developed as an illustrative case so that additional figures and measures can be identified. The figures and measures that were identified for the first-order model are still applicable to this more complex model. The overall FOM related to the costs or benefits of a decision system is also the single FOM for evaluators.

For the configuration that will be developed, it is assumed that the evaluator now has access to internal states in the decision system (otherwise the evaluation would be identical to the first-order model). An alternative and equivalent model configuration is one in which the internal states of the decision systems are still hidden, but the decision systems' actuators can write the internal states into the environment for the evaluator to read and process.

The same mathematical foundation that was used in the first-order model will also be used in the second-order model. If the model is assumed to form a state machine, the evaluator can be considered able to either exactly or probabilistically measure the world state. Between the two model configurations of an evaluator with perfect knowledge and an evaluator with probabilistic knowledge, the evaluator with perfect knowledge of the world state is the easier to process. Inexact knowledge of the world state leads the probabilistic evaluator to adopt a PDF to estimate the likely states of the environment. A necessary requirement for productive evaluators is that they have a better estimate of the world state than the decision systems under test.

In this model configuration, the evaluator makes the assumption that the decision system is a probabilistic decision system and that it uses probability theory to make its decisions. The evaluator will view most of the internal process in the decision systems as generating PDFs (or their parametric equivalent) for the measurements, features, and world state. Even if the decision system does not actually use probability theory as a foundation for its decision process, the evaluator can map that process to a probabilistic foundation for the purposes of evaluation.

Statistical decision theory can be adopted as a foundation for the decision-process step, meaning that the simplest output of the step is a specific decision that commands the actuator. If the decision-process step produced a series of costs for possible decisions, the actuator would activate the decision associated with the lowest cost. The cost calculation for statistical decision theory requires an estimate of the cost $C(d, E)$ for all possible decisions, d , given all possible states of the environment, E . The matrix-vector product of the cost matrix with the probability vector of the possible environmental states results in a vector of costs for the decisions, as shown in Equation (4).

$$C(d) = \sum_{d, E} C(d, E)P(E). \quad (4)$$

For evaluation purposes, the evaluator can compare its PDFs with those generated by the decision system. In general, a good decision system's PDFs will agree more often with the evaluator's estimate of the truth than would the PDFs of a bad decision system. A probabilistic agreement function that provides a numerical comparison between PDFs is

$$D_B = \sum_x P_E(x)P_D(x), \quad (5)$$

where D_B is the numerical measure, $P_E(x)$ is the PDF from the evaluator, $P_D(x)$ is the PDF from the decision system, and x is the measurement, feature, or world state space or subspace of interest to the evaluator. This function measures the agreement between the two PDFs when the true value of x is not known. If the evaluator knows the true value of x , then $P_E(x)$ is a delta function, $\delta(x - x_r)$, and in this case,

$$D_B = P_D(x_t). \quad (6)$$

Now that the complexity of the model supports multiple state spaces between processes, information theory can be used to evaluate the performance of the internal processes, treating internal decision processes as analogous to communications channels. In the simplest model, the mathematical transform between sensors and actuators looped back into the environment and did not have any internal structure to model as a communications channel. Information theory models channels between a unique source and destination.

Shannon information, defined in Equation (7), provides a measure of the uncertainty (or entropy) of a PDF such as those generated by processes within the decision system.

$$H(x) = -\sum_x P(x) \ln(P(x)). \quad (7)$$

Smaller entropies correspond to more accurate estimates of the state. This measure only relates to precision; it does not reflect whether an estimate is correct. The PDF could be completely incorrect but still have a small value for its entropy.

Another information theoretic measure, relative entropy, or Kullback-Leibler distance, can be used as another measure of the similarity of two PDFs.

$$D(P_E \parallel P_D) = \sum_x P_E(x) \ln(P_E(x)) - \sum_x P_E(x) \ln(P_D(x)). \quad (8)$$

Relative entropy is always non-negative and is zero if and only if $P_E(x) = P_D(x)$. The convention is that $0 \ln(0) = 0$ and $-P_E(x) \ln(0) = \infty$. This measure is not a metric distance because it is not symmetric. A symmetric version of the Kullback-Leibler distance, D_S , is often used to avoid specifying the preferred PDF.

$$D_S(P_E \parallel P_D) = D(P_E \parallel P_D) + D(P_D \parallel P_E). \quad (9)$$

Because the internal processes in the decision system are analogous to communications channels, the information loss that occurs due to inefficiencies in the transformation processes can be measured. Mutual information provides a measure of the common information content on both sides of the process,

$$I(X; Y) = -\sum_x P(x) \ln(P(x)) + \sum_{x,y} P(x, y) \ln(P(x|y)). \quad (10)$$

The evaluation of this measure requires a PDF, $P(x, y)$, for the combined space of both the space x and the space y . The conditional PDF, $P(x|y)$, is also required, but can be derived from $P(x, y)$.

An evaluator will need to collect measurements of both spaces and generate a combined estimation of the PDF. For the second-order model, an evaluator could compare PDFs that are generated anywhere along the chain of processing to identify inefficiencies.

4.3 THE THIRD-ORDER DECISION MODEL

The additional complexity of the second-order decision model has increased the power of the model but is still not sufficient to adequately describe enterprises built on service-oriented architectures. The model needs to be extended to capture the key attribute of SOAs; these architectures consist of a set of distributed, loosely connected systems. The final extension is made to the model by linking multiple second-order models together into a single distributed system, as shown Figure 3. This third-order model now contains multiple decision systems that interact with each other within a common environment and that communicate with each other by passing data through the environment. In this model, each second-order system is striving to change the environment to a more desirable state, or at least one that it views as desirable.

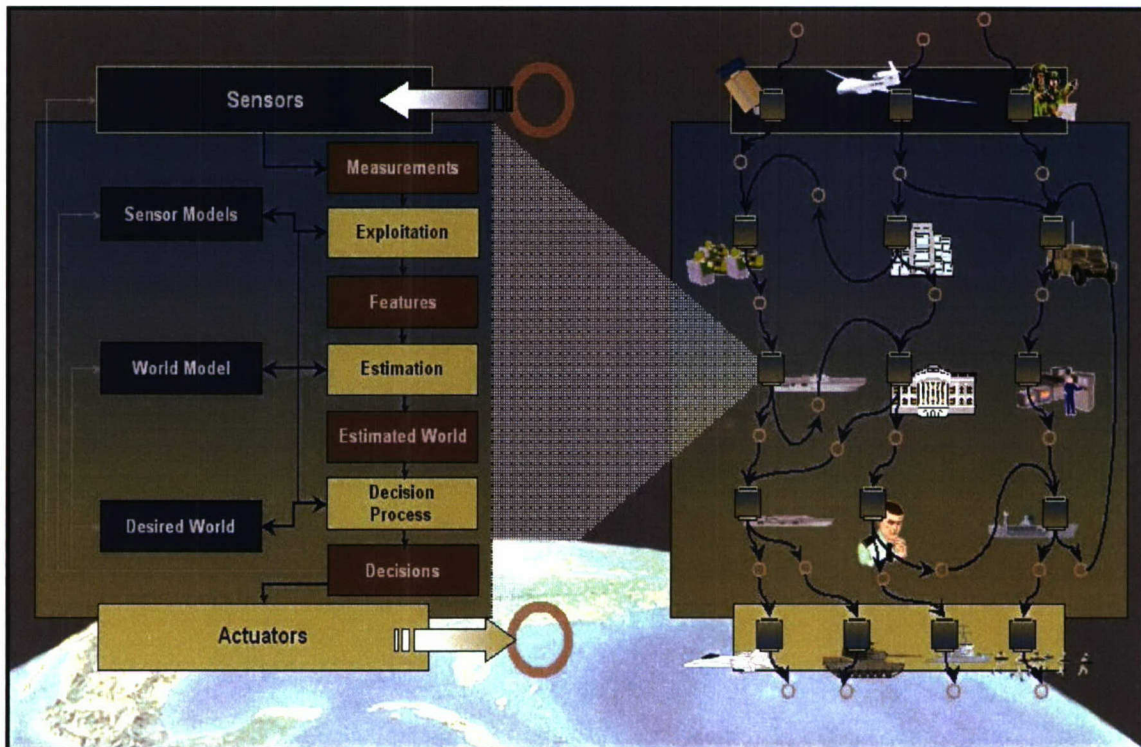


Figure 3. The third-order decision model.

An interesting feature of the extension to the third-order model is that each of the component systems performs one of the information transformations identified in the second-order model: measurements to features, features to world state estimates, or state estimates to decisions. The second-order decision systems thus form a cooperative higher-level organization that emulates the functions of the second-order decision model in a distributed system.

This emulation leads to a profound difference between the two decision models of Figures 2 and 3: a component of the third-order decision system has the potential to do more than its assigned primary role in the higher-level decision system. It may perform all the transformations of the fundamental decision system in addition to its primary role. It can make measurements, generate features and environmental state estimates, pursue goals, and learn. This secondary processing might not be under the direct control of the higher-level decision system and means that a component could be pursuing goals not in the best interests of the higher-level organization.

An interesting difference between the second- and third-order models is that in the second-order model, the sensor, world, and desired world models were globally available to the decision system; in the third-order model, these supporting models are not assumed to be global. In general, distributed decision systems will not have a unified set of sensor, world, or desired world models, but can allow each component system to define these models locally. This may result in the component systems of the distributed system producing different, and potentially inconsistent, data products and PDFs due to differences in their local models.

In retrospect, the internal sensor, world, and desired-state models of the second-order decision system can be redefined so that the models are actualized as stored configuration data. The transformation processes in the second-order model are then assumed to be sufficiently powerful and flexible enough to read the configuration data, set up their internal processors, and execute any reasonable sensor, world, or desired-state model.

Given that a common design goal for the construction of a third-order system will be to emulate the processing of the second-order model, it is possible that this intention may go awry and the third-order model will not perform all of the transformations of the second-order model. Some measurements, features, and environmental state estimates may not exist because the system never calculates them and can increase the challenge of analyzing a poorly designed system.

Because the components of the third-order systems can be connected in a myriad of configurations, it is possible that multiple components will process the same data. These components may represent and interpret the same data in different ways. These differences in interpretation and usage can be valuable in that different components will have different sets of subgoals that are coordinated for the common good of the organization. On the other hand, these differences can also cause problems of misinterpretation and miscommunication when components fail to recognize that other components have adopted different representations for the data and have processed the data to achieve different subgoals.

Although the third-order model is constructed from second-order decision systems, not all second-order components are required to instantiate all internal transformations. Uninstantiated internal transformations can be interpreted as transformations that directly pass data to the next stage without conversion (mathematically, these are identity functions) so that the component only performs a subset of the processing that the second-order decision system can perform. At the furthest extreme, the internal processing of a component can be so simple that it looks like a first-order decision system that directly transforms sensor data to actuator actions. This potential heterogeneity in system components means that it may be more difficult to model some enterprises because of the inability to model the components as simple, identical agents. This also means that the analysis of relative component performance may be more difficult because not every component will have internal data or PDFs that evaluators can compare with the truth or the internal data or PDFs of other components.

The evaluation process may be more difficult not only because of limited access to the internal data of the decision system components but because the components that have been assumed to be second-order decision systems may be, in themselves, collections of multiple, loosely coupled decision systems. These lower levels in the decision process may be hidden from an evaluator and may make analysis even more difficult. In the worst cases, there can even be couplings between different levels in the hierarchy of decision systems.

Because the components of the third-order model may be widely distributed, problems may arise from the transfer of data between components. Data transfers may not always be performed successfully or data may be degraded through interactions with the environment. Communications-related issues will become a major focus of the evaluation process, with information theory providing the natural mathematical foundation for understanding these issues.

4.3.1 Evaluation of Third-Order Decision Models

There are many different options for the evaluation of third-order models because the systems can be so rich and varied. All the measures that were proposed for the evaluation of the second-order model can be applied to the third-order systems as a whole, as well as to the system components. The number of PDF comparisons that can be made increases from just between those of the evaluator and a single decision system to comparisons between any or all of the components of the system (including the evaluator). For example, the probability distance function can be used to compare two PDFs generated by two different components (or that have been transferred between components with possible losses during transmission):

$$D_B = \sum_x P_{Da}(x)P_{Db}(x), \quad (11)$$

where $P_{Da}(x)$ is a PDF from component a and $P_{Db}(x)$ is a PDF from component b . The PDFs must be for the same feature space x .

Shannon information, or entropy, can be used to evaluate two PDFs on the same features space,

$$H(x) = - \sum_x P(P_{Da}(x), P_{Db}(x)) \ln(P(P_{Da}(x), P_{Db}(x))). \quad (12)$$

Estimating the second order probability distributions of first-order PDFs is much more difficult than estimating the first-order probabilities, but it is conceptually possible. The entropy for this second-order PDF would be a measure of the similarity between the two component PDFs.

More that two PDFs can be compared for overall consistency with the chain rule,

$$D_B = \sum_x \prod_{i=1}^n P_i(x). \quad (13)$$

Shannon information, Kullback-Leibler distance, and mutual information can be used to evaluate pairs of distributions and collections of distributions through chain rules, as well. The chain rule for Shannon information is

$$H(x_1, x_2, \dots, x_n) = \sum_{i=1}^n H(x_i | x_{i-1}, \dots, x_1) \quad (14)$$

using conditional entropy to estimate the subsequent contributions,

$$H(x_2 | x_1) = - \sum_{x,y} p(x,y) \ln(p(y | x)). \quad (15)$$

The chain rule for the Kullback-Leibler distances is

$$D(p(x,y) || q(x,y)) = D(p(x,y) || q(x,y)) + D(p(y | x) || q(y | x)). \quad (16)$$

The chain rule for mutual information is

$$I(X_1, X_2, \dots, X_n; Y) = \sum_{i=1}^n I(X_i; Y | X_{i-1}, \dots, X_1). \quad (17)$$

If the appropriate conditional density functions can be constructed, overall measures for such things as PDFs associated with common operational pictures can be estimated.

FOMs and measures of effectiveness still require value assignments in order to estimate the expected costs or benefits that are generated from the measures of performance. The evaluations, like the measures of performance, can be carried out on the global system and subsets of the system. Component level analyses are valuable to locate inefficiencies and bottlenecks in the larger system.

5. A DECISION MODEL FOR ISR ENTERPRISES: A SIMPLE DEMONSTRATION

The military C4ISR enterprise can be easily mapped to the third-order decision model. Military units in the C4ISR enterprise are often assigned unique roles that correspond to the transformations in the second-order decision model, which include sensing the battlespace, reporting measurements, developing situational awareness, making decisions, and issuing commands to control the military organization. The C4ISR organization is divided into two major components: an ISR community that processes measured data to report features and a command-and-control community that processes features to develop an operational picture of the battlespace and determine appropriate courses of action.

The FOMs, measures of performance, and measures of effectiveness for all three levels of the models can be used to evaluate the C4ISR enterprise. As the enterprise is decomposed into the ISR and command-and-control enterprises, evaluation of these specialized enterprises becomes problematic with regard to FOMs and measures of performance because the entire enterprise is required to estimate these parameters. FOMs can only be estimated for the subsystems if these subsystems can be viewed as independent third-order decision systems, with their own goals and values for the environmental states and with their own actuators used to achieve the desired world states. For example, it might be an interesting analysis to view the ISR enterprise as an independent, complete, third-order decision system that must achieve its goals by manipulating the command-and-control enterprise into engaging its actuators to transition the environment to a state that is desirable to the ISR enterprise. It is unlikely that this Machiavellian scenario would be a popular state of affairs with any senior commander that oversees a C4ISR enterprise. Given this, a global FOM for the ISR component cannot be developed without knowledge of the value that the overall C4ISR system assigns to the possible world states. In addition, the probabilities that different environmental states will be achieved cannot be estimated without accounting for the performance of the entire organization. Given that assessments of the ISR enterprise, as well as the command-and-control enterprise, will still be conducted without information on the rest of the enterprise, the assessments of these components will be primarily restricted to measures of performance and their differences and ratios.

The DCGS-N program office has indicated there is interest in using the results of this analysis to assess SOAs. The DCGS-N application is an ISR Multi-INT system that primarily generates features for the command and control portion of the C4ISR enterprise. The added value of DCGS-N over the current system is through the reduced time to find, process, and report data. Ideally, this should improve the quality of features (PDFs) that are reported to the command-and-control enterprise.

The services that DCGS-N will provide through its SOA can be mapped to the third-order model. These services include metadata repository services (data storage), data discovery services, data access services, workflow services, and data analysis (translation) services. The data storage, discovery and access services, as well as workflow services, support the formation and reformation of the links between the component systems' inputs and outputs, represented in Figure 3 as the arrows between the decision

subsystems and the circles in the environment. The data translation services in SOAs are often automated services that translate raw data to more usable formats for other components in the decision system—such as human operators. These services would be represented as decision subsystem components in Figure 3.

In order to conduct an initial assessment of the applicability of the probabilistic and information theoretic measures to the evaluation of the ISR enterprise, a simple simulation was developed. An active issue in the C4ISR community was chosen as the basis for the simulation. Given that many evaluations of ISR performance will be conducted in isolation from the full C4ISR community, the issue was restricted to one involving measures of performance instead of measures of effectiveness and figures of merit. A simulation of a distributed system that develops a COP was chosen as the model for the investigation. In the third-order model, the COP is synonymous with the world model. It is interesting to note that the Navy has the COP processing embedded in a C2 application instead of in an ISR application. The model that has been developed here would more naturally assign the COP processing to the ISR enterprise.

The third-order model of the C4ISR enterprise suggests that additional problems can still be encountered, even when an accurate COP is shared between the components in the enterprise. Even with an accurate, widely distributed COP, system performance may still be below optimal levels because the goals in the subsystems are not aligned. Military organizations have been aware of the need to align goals across the organization for millennia. Solutions have included the development of the chain of command, military uniform codes of justice, and training and education to enforce the adoption of consistent goals throughout the organization.

The second-order model shows that a component in a third-order system requires a series of conversions for goals to be altered, starting with sensing (receiving new orders), then feature extraction, then world state update, and finally a decision process that updates the internal goals of the system to reflect the new orders. Clearly, the third-order model can be used to analyze problems with command-and-control systems. The chain of transformations that the model defines can be applied to study command-and-control problems caused by such factors as information loss, misinterpretation, and even rogue units that refuse to follow orders because their own goals are more beneficial.

In order to demonstrate the utility of the proposed effectiveness measures, a simple experiment was developed in a simulation environment. In the experiment, a set of heterogeneous decision systems attempts to correctly detect and classify several objects through a series of noisy observations. Although simple, this experiment is grounded in tasks central to the ISR enterprise. The information theoretic metrics suggested above were implemented in order to compare a variety of enterprise architectures, with encouraging results.

The environment for the simulated experiment consists of a 5×5 grid world. Each grid location contains at most one object of interest to the ISR enterprise. In reality, these could be vehicle types or combatants, but in the simulation they are abstractly referred to as circles, squares, and triangles. Thus, each cell contains a circle, a square, a triangle, or none of the above.

Additionally, five sensors, each with a unique sensing modality, exist on the grid. The characteristics of the various sensors with regard to detection and discrimination are summarized in Tables 3 and 4. Each sensor is located in a specific cell and has the opportunity to move to an adjacent cell at each time step. The simulation also allowed a limited amount of communication between each of the sensors. The sensors are decision systems in the sense that they must decide: 1) which cells within the grid to explore, and 2) whether and how to share the observed results with other sensors in the enterprise.

Table 3
Sensor Detection and Discrimination Strengths

	Sensor A	Sensor B	Sensor C	Sensor D	Sensor E
O	1	3	5	1	1
□	1	3	1	5	1
Δ	1	3	1	1	5
None	0	0	0	0	0

Table 4
Sensor Observation Extent
(Edge length of the observable subgrid)

	Sensor A	Sensor B	Sensor C	Sensor D	Sensor E
Extent	5	3	1	1	1

As an example, if sensor B were located at a specific cell in the 5×5 grid (denoted by the large “B” in Figure 4), it would be able to observe all nine cells in the 3×3 subgrid centered on its location. For each observed cell that contains a circle, square, or triangle, the resultant measurement would be $3 + \eta$ where η is an additive noise term. If the cell does not contain any of the objects, the resultant measurement would be $0 + \eta$. (All noise terms are drawn from zero mean, unit standard deviation Gaussian distributions.) This scenario is demonstrated in Figure 4.

It is significant to note that sensor B has no ability to discriminate between a circle, square, or triangle; all return statistically identical measurements. This means that without data sharing from other sensors in the enterprise it is impossible for sensor B to discriminate between all the objects of interest. The same is true of each of the other four sensors.

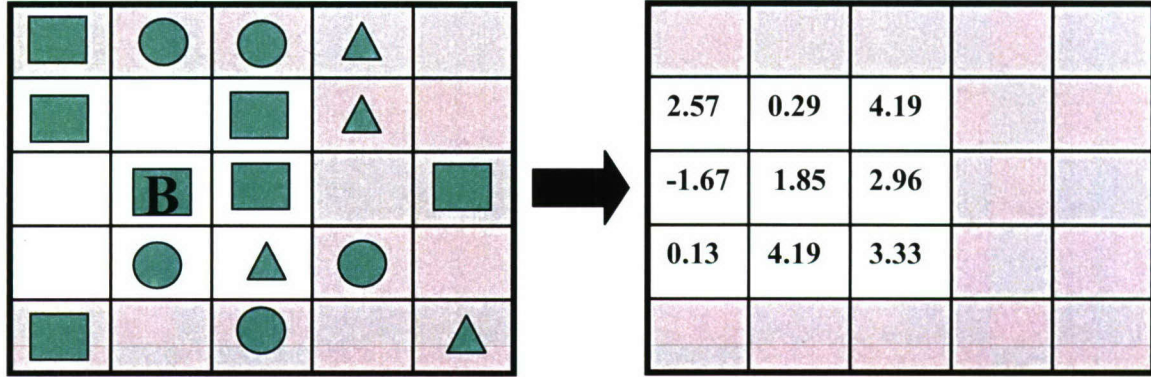


Figure 4. Demonstration of sensing capability for sensor B.

In the experiment, five different enterprise architectures (described in the paragraphs below) were run through a series of 300 Monte Carlo trials each. The information theoretic measures of entropy, Kullback-Leibler (KL) distance, and Bayesian agreement with truth (Bayesian distance) were collected and aggregated across the trials. Figures 5–9 demonstrate the power of these metrics in evaluating the performance of the various architectures.

The first architecture examined was termed the “unlimited communication” architecture. In this architecture, no limitations on communication bandwidth were made, and, therefore, the ability of the enterprise to share and assimilate knowledge was limited solely by the rate of data acquisition. The sensors used Bayesian logic to fuse all measurements from all sensors into their local world models. This idealized architecture provides an upper bound on the levels of expectable performance. The statistical results, shown in Figure 5, demonstrate a rapid convergence to absolute certainty (measured by Shannon entropy), complete consistency (measured by KL distance), and perfect correctness (measured by Bayesian agreement with truth). The small amount of initial inconsistency is due to latency effects, and the convergence rates of the other two metrics are due to the noise characteristics of the environment.

In the second architecture, called the “no communication” architecture, there was no communication between any of the decision systems. This worst-case scenario was used to establish a lower baseline on the potential performance of the other architectures. In this architecture the sensors used Bayesian logic to update their local world models, but did not communicate any gained knowledge with any other sensors in the enterprise. The results, shown in Figure 6, are that the world models of the various sensors do not achieve a high degree of certainty, consistency, or correctness.

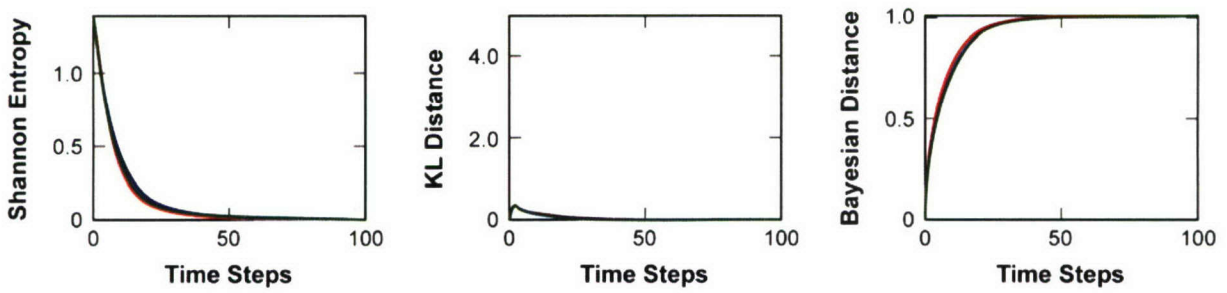


Figure 5. Average Shannon entropy, KL distance, and Bayesian agreement with truth for each of the five sensors when using the “unlimited communication” architecture.

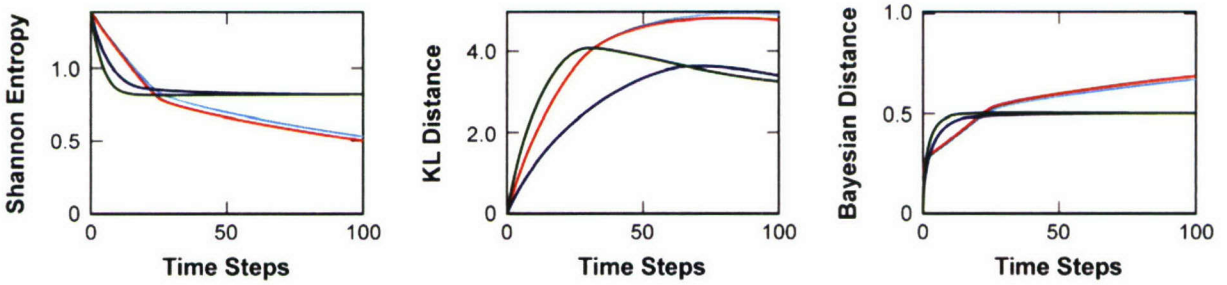


Figure 6. Average Shannon entropy, KL distance, and Bayesian agreement with truth for each of the five sensors when using the “no communication” architecture.

The “unlimited communication” architecture and the “no communication” architecture provide upper and lower bounds, respectively, on the reasonable performance of communication constrained enterprise architectures.

The final three architectures investigated all assume a uniform, fixed bandwidth constraint on node-to-node communication. In the first architecture, called the “blind push” architecture, all sensors attempt to push all measurements to all other sensors. Due to the communication constraints, this results in an earliest-first queue of measurements at each node. The performance results of this architecture are shown in Figure 7. Because of the absence of a prioritization mechanism, important measurements got jammed in the communications bottlenecks and all the performance measures suffered accordingly. While not as bad as the “no communication” case, there is significant performance degradation from the optimal “unlimited communication” case.

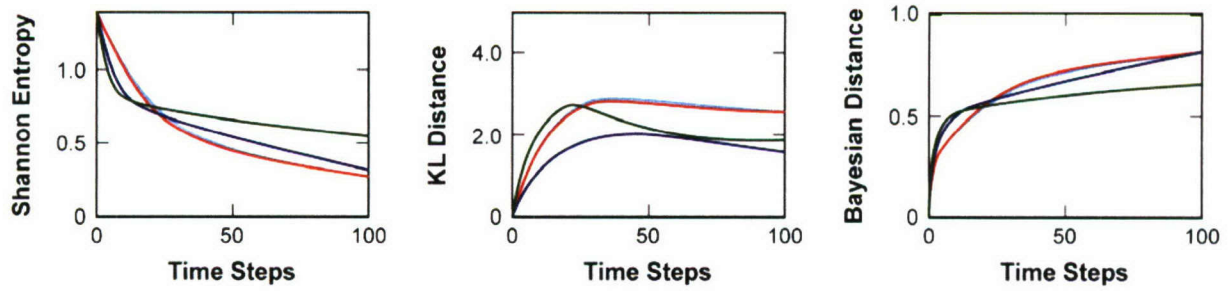


Figure 7. Average Shannon entropy, KL distance, and Bayesian agreement with truth for each of the five sensors when using the “blind push” architecture.

The second fixed bandwidth architecture was termed the “blind pull” architecture. In this architecture sensors were allowed to broadcast a request (which took a portion of the bandwidth) at each time step. These requests enabled the other sensor nodes to prioritize the measurements they pushed to that sensor node. However, because the broadcast request consumed some bandwidth, fewer measurements could be communicated. The results, shown in Figure 8, show that despite receiving fewer measurements from other sensors, the rudimentary prioritization mechanism improved enterprise performance over the “blind push” case.

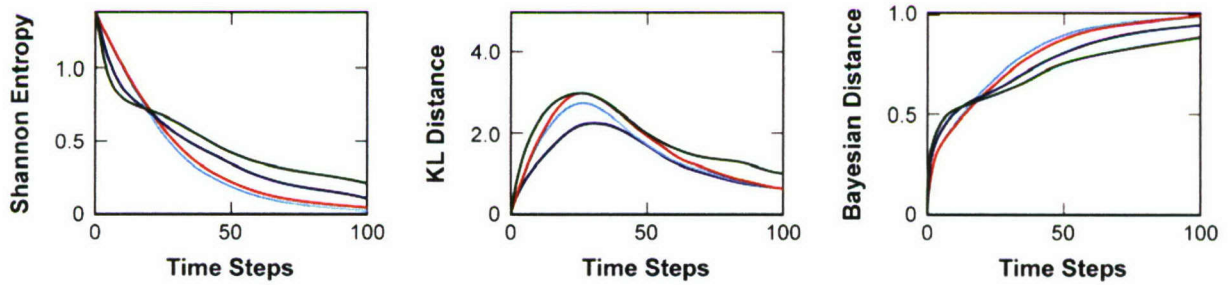


Figure 8. Average Shannon entropy, KL distance, and Bayesian agreement with truth for each of the five sensors when using the “blind pull” architecture.

The final architecture investigated was called the “informed pull” architecture. In this variant, each sensor node broadcast its position information at each time step. This can be thought of as the metadata for that sensor’s measurement at that time step. Broadcasting the metadata consumed a portion of the bandwidth, resulting in fewer measurements being communicated than in either the “blind push” or the “blind pull” architecture. However, the metadata enabled the other sensors to direct requests for specific

data products, creating an efficient prioritization scheme. The performance results are shown in Figure 9. There is a dramatic increase in the ability of the enterprise to maintain consistency among the distributed nodes, to quickly converge to a high level of certainty, and to correctly classify all the objects of interest.

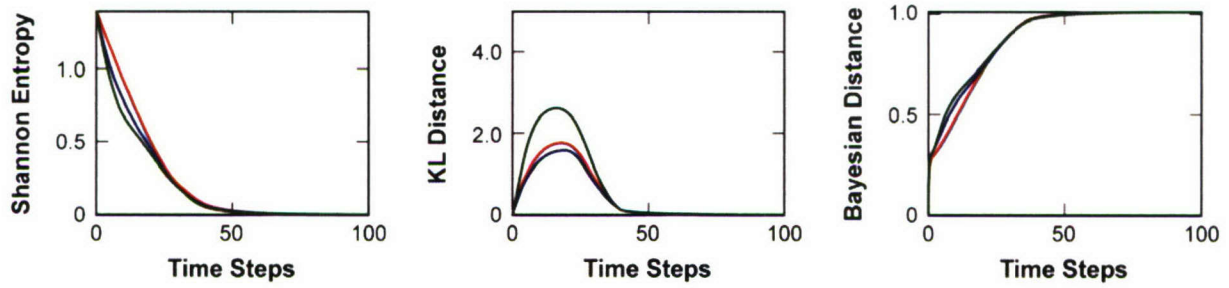


Figure 9. Average Shannon entropy, KL distance, and Bayesian agreement with truth for each of the five sensors when using the “informed pull” architecture.

6. CONCLUSIONS

Defining a single FOM, or a suite of FOMs, enables the system developer to make informed decisions regarding design trade-offs. With regard to the ISR enterprise, the proposed FOMs provide the ability to compare the relative effects of a multitude of potential measures of performance that may have different units and obscure relationships to one another. Comparisons can be made between improving sensor capabilities versus increasing bandwidth, or adding new processors versus improved training and manning. This evaluation capability is essential to effective enterprise development and design.

In the course of this study, a series of conceptual decision models were constructed with increasing levels of complexity to form the basis for characterizing distributed multi-INT and ISR enterprises. Because these decision models are based on a strong statistical foundation, they provide natural, quantitative measures for the characterization of enterprise performance. Probability theory and information theory were selected as the mathematical foundations because of their maturity, ability to account for uncertainty, and capability for modeling the information flow. The models not only provide a mathematical foundation for evaluating enterprise performance, but a conceptual foundation for understanding the underlying mechanisms by which enterprises accomplish their performance objectives.

The conceptual models show that the truly important FOMs require an evaluation of the entire decision system: from sensors, through decision processes, to actuators with regard to the costs or benefits associated with the systems running within their environment. Most of the FOMs that have been adopted for prior defense systems have actually been either sensor or actuator measures of performance; in very few cases were they decision process measures. The sensor measures of performance relate to the ability of the sensors to measure the environment; the actuator measures of performance relate to the ability to change the environment; and the decision process measures of performance relate to the ability to understand the environment and decide how to change it. With the “stove-piped” systems of the past, the sensor and actuator measures were more pertinent, but as ISR systems are transformed into distributed enterprises, the decision process measures will gain in importance.

A “toy” simulation was developed to demonstrate the use of probabilistic and information theoretic measures to assess a distributed multi-INT ISR enterprise. The simulation was constructed to demonstrate how the proposed information theoretic FOMs could be used for system development, in this case through evaluating the relative merit of several different information-sharing architectures. The similarity and accuracy of the sensors’ operational pictures were evaluated with Shannon information, Kullback-Leibler (KL) statistical distance, and Bayesian agreement measures. The results of the simulation show that these measures provide an accurate quantitative assessment. They agree well with what an expert’s intuition would predict with regard to how the common operations pictures would change with changes in the communications architectures.

This study provides an initial step in the development of performance assessment measures for multi-INT and ISR enterprises. Although the tone is optimistic for completing these assessments for real

systems, the researchers recognize that adapting these measures to real systems remains an open problem. Further research is required to identify how to construct specific models of real systems, determine the information, costs, and benefits of true interest to the end-users, collect the relevant data, and perform the assessments.

APPENDIX A

METRICS FOR OTHER MODEL ATTRIBUTES

In the process of conducting the literature search, a number of different papers were found that proposed attributes of information quality. Some lists were quite extensive. Quantitative measures were ad hoc for the most papers, if specified at all. This appendix suggests probabilistic and information theoretic measures that can be related to some of the attributes.

A.1 ACCURACY

The scientific community recognizes two terms: accuracy and precision. This distinction will be used in the discussion of other attributes although common usage usually neglects to make this distinction. Precision will usually be the dominant measure of interest for decision systems because the truth is not known, and accuracy will be the dominant measure of interest for evaluators because the truth is usually known.

Precision can be defined to be inversely proportional to the standard deviation, σ , of a PDF. The functions,

$$\mu = \sum_x xP(x), \quad (18)$$

$$\sigma^2 = \sum_x (x - \mu)^2 P(x), \quad (19)$$

estimate the mean, μ , and standard deviation of a PDF. The standard deviation is considered to be inversely proportional to precision. A measure of accuracy, σ_A , can be estimated through a similar equation, in which the mean is replaced with the true state, μ_T :

$$\sigma_A^2 = \sum_x (x - \mu_T)^2 P(x). \quad (20)$$

Shannon information, or entropy, of a PDF can be used as a measure of precision as well. The greater the entropy, the more uncertain and imprecise is the PDF.

A.2 CONSISTENCY

In some respects, this is a secondary measure. Consistency is most valuable when there is consistency with the truth. A lesser level of value might be assigned to consistency when inconsistency

leads to conflicting decisions throughout the community that are more costly than consistent, but wrong, decisions.

Probabilistic measures, such as the probabilistic agreement measure, and information theoretic measures, such as the KL distance, provide measures of consistency between PDFs.

A.3 SECURITY

Security is an attribute that is beyond the level of analysis for this study. Information theoretic measures could possibly be used to measure the mutual information between subsystems that are or are not to have access to certain information. If two subsystems—one of which has access to secure data and the other of which is denied access to the data—have high mutual information between their measurements, features, or world subspaces, then either the secure data are available to the denied system or the secure data are not contributing unique information to the secure system.

Network analysis and connectivity diagrams provide other means beyond probabilistic techniques for evaluating the security of data. Probabilistic techniques might provide additional methods for the security assurance community.

A.4 LATENCY AND TIMELINESS

This attribute directly relates to measures of precision and accuracy because of its specific impact on the PDF—through the time variability of the PDF. If it is accepted that the PDFs include a dimension of time or that appropriate models exist to propagate a PDF to any point in time, then the equations for measuring precision and accuracy are directly influenced by effects due to latency. The time evolution of the PDF usually decreases the precision of the PDF. The decreased precision of the PDFs leads to decreased confidence in the decisions that the decision system makes, which is quantified by increases in risk.

The increase in Shannon information attributable to the latency L can be estimated with the equation

$$\begin{aligned} \Delta H(x) = & - \int_t^{t+L} P(E | x_1, \dots, x_N, i, t) \ln(P(E | x_1, \dots, x_N, i, t)) dt \\ & + \int_t^{t+L} P(E | x_1, \dots, x_{N+1}, i, t) \ln(P(E | x_1, \dots, x_{N+1}, i, t)) dt \end{aligned} \quad (21)$$

where $P(E | x_1, \dots, x_N, i, t)$ is the time-dependent PDF estimated at time $t + L$ using measurements x_1, \dots, x_N and $P(E | x_1, \dots, x_{N+1}, i, t)$ is the time-dependent PDF estimated at time t using measurements x_1, \dots, x_{N+1} , with no latency.

If the decision system uses cost functions to make decisions, the increased risk that the system faces can be evaluated with the function,

$$\langle \Delta C(d, E) \rangle = \frac{\int_t^{t+L} \sum_E C(d, E) [P(E | x_1, \dots, x_N, i, t) - P(E | x_1, \dots, x_{N+1}, t)] dt}{(t + L - t)} \quad (22)$$

A.5 COMPLETENESS

Completeness can be related to the number of independent PDF axes with informative subspace PDFs, with the collection of axes defining multidimensional PDFs. For a multidimensional PDF, there may be many axes on which the marginal PDF for some subspace of features is minimally informative or insufficiently informative. The relative numbers of informative and uninformative axes provide an analytical measure of completeness. Extracting out the subspace PDFs and evaluating the Shannon entropy of the subspace PDFs can measure the information within the axes. If the entropy is too high, the information axes can be considered incomplete.

A.6 CONCISENESS

Conciseness can be related to the compression of information. Measures of conciseness are beyond the scope of this study. Information theoretic approaches, as well as approaches from coding theory and complexity theory, can be used to estimate the conciseness of the representation of the PDF. Compression reduces the fraction of noninformative data contained in a data format.

A.7 RELIABILITY

This attribute refers to the accuracy of the data provided by one subsystem to another. A reliable subsystem (a reliable source) is correct more often than is an unreliable system. The probabilistic and information theoretic measures for accuracy can be used to measure reliability. The reliability measures may need to account for latency, information loss between subsystems, and missing messages. More complex decision subsystems could conceivably attempt to estimate the reliability of its data providers with a secondary decision process (see A.15, “Believability”). Additional analysis of reliability measures is beyond the scope of this study.

A.8 ACCESSIBILITY

This is a secondary attribute with respect to the model. The connectivity between subsystems, and the costs and delays for the transmission of data between subsystems would influence estimates of accessibility. Measures of accessibility are beyond the scope of this study.

A.9 AVAILABILITY

This attribute is related to accessibility and probably refers to a zero-one type of decision: either the information generated by a subsystem can be transported to the sensors of another subsystem or it cannot. The connectivity of the communications links between sensors would provide for the estimates of the availability of data throughout the environment. Measures of availability are beyond the scope of this study.

A.10 OBJECTIVITY

This attribute can deal with the biases and prejudices of subsystems. Probabilistic measures can be used to estimate biases, such as estimates for skew:

$$\gamma^3 = \sum_x (x - \mu)^3 P(x) / \sigma^{3/2} , \quad (23)$$

where this is only one of many different definitions for skew. The measure of the intersection of truth with the PDF function indirectly reflects the effect of bias and prejudice on the estimate. Covariance estimates with respect to the PDF and truth can indirectly indicate the presence of biases and prejudices.

Trend analysis can uncover biases and prejudices in a system by comparison of the measures like skew, covariance, entropy, and the KL distance. Additional study is required to fully evaluate all possible quantitative measures of objectivity and probably requires multiple classes to fully quantify all aspects of this attribute.

A.11 RELEVANCE

Mutual information provides a measure of relevance. Features that have high measures of mutual information with the value of world states or decisions are more relevant. Decision systems that learn are most likely continually assessing the relevance of the sensor measurements against goals and actions. Once measurements are identified as irrelevant, they are ignored. Measures of mutual information between measurements, goals, actions, and resulting environmental states provide a means to quantify relevance.

A.12 USABILITY

Usability relates to the ability of the subsystems to convert data in one format or representation into other formats, representations, or state spaces that are more easily manipulated or processed through the remaining stages of the decision process.

Being able to measure subspaces of the world through the sensors does not mean that the subsystem will be able to extract information or meaning from those measurements. Coding theory, complexity theory, and other information theoretic methods may be able to contribute quantitative measures of usability. The identification of these measures is beyond the scope of this study.

A.13 UNDERSTANDABILITY

This attribute makes reference to the lack of ambiguity and ease of comprehension. In terms of mathematical mappings, it could refer to the number of peaks that the output PDF might have, based upon a given input PDF. More peaks would mean that there are more interpretations that could be selected by the decision-processing step. It could also refer to the ease that a decision system has in mapping PDFs in one coordinate frame to a coordinate frame with more relevance to the system. Mutual information would be a quantitative measure of the information loss resulting from the transformation process. Additional analysis is required to fully develop quantitative measures of understandability and is beyond the scope of this study.

A.14 AMOUNT OF DATA

The amount of data is a straightforward calculation; the impact of this attribute is more difficult to surmise. The information-quality community adds assessments as to whether the appropriate amount of data was available. Both too much and too little data can be a bad thing. It is possible that the accuracy of PDFs could be connected to the amount of data, in that more data should produce more accurate PDFs with a diminishing rate of return as more data are integrated into the PDF estimate. Cost-benefit analyses might be required to assess the value of each additional data point and look for the point of diminishing return or the transition from a positive benefit to a cost. Additional analysis of suitable quantitative measures is beyond the scope of this study.

A.15 BELIEVABILITY

This is a judgment by a consumer of the data as to whether it is true and credible, given the prior information that the consumer has acquired. A decision subsystem could execute secondary decision processes to analyze past performance of data providers and develop a ranking of the believability of the provider. This attribute is similar to reliability. The distinction appears to be that evaluators would tend to be interested in reliability, and decision subsystems would tend to be interested in believability. Similar quantitative measures might be applicable to both attributes of reliability and believability. Additional analysis to identify possible quantitative measures is beyond the scope of this study.

A.16 NAVIGATION

This sensor performance attribute relates to the ease of finding and linking to data. This is a very important performance measure for SOAs, but is a fourth-order measure at best, first-order being FOMs, second-order being measures of effectiveness, third-order being measures of performance directly relevant to PDFs, and fourth-order being parameters that are component variables that can be combined with other fourth-order measures to define a PDF. Other theories, such as those involving networks and connectivity measures, might provide important quantitative measures. Additional analysis to identify possible measures is beyond the scope of this study.

A.17 REPUTATION

This is a measure of how much trust a collection of data consumers has in a producer. This consensus building on the part of the data consumers could be done through the collection of quantitative measures for any set of attributes listed in this section, especially reliability, objectivity, and believability. Reputation is the attribute that inexperienced data consumers use to rate the data from producers that they have not interacted with before. Given the complexities of this attribute, the identification of possible quantitative measures is beyond the scope of this study.

A.18 USEFULNESS

This attribute is related to the extent that received data is useful for making decisions. One measure could be a comparison of the relative change in the information content of a consumer's PDF before and after the data have been used to update the PDF. Small changes in Shannon information would indicate that the data were not very useful. The KL distance between the two PDFs would provide another measure of the usefulness of the new data. In this case, more useful data will cause a larger increase in the KL distance between the prior and updated PDFs than will less useful data.

If the true state is known, the probabilistic comparison of the two distributions with truth could be calculated. If the updated PDF has a larger Bayesian correctness measure, it is closer to the truth than the PDF was before the update.

A.19 EFFICIENCY

This attribute relates to the energy that a decision subsystem is required to expend in order to put the received data to direct use. Traditionally, quantitative attributes, such as the percentage of useful work that can be extracted for a given input of energy or energy expended per unit of information, have been defined for systems. Although this is an important attribute with a number of important quantitative measures waiting to be defined, additional analysis is beyond the scope of this study.

A.20 VALUE-ADDED

This is a measure of the extent that provided information reduces costs or increases benefits to an organization. Clearly, cost-benefit analyses are part of possible quantitative measures for this attribute. In most cases, this would be a relative comparison of the value that different subsystems add to the collective higher-level decision system. Estimates can include not only the decision subsystems, but also the sensors and actuators. Additional analysis is beyond the scope of this study.

REFERENCES

1. P.A. Pomianowski, R.L. Delanoy, J.R. Kurz, and G.R. Condon "Silent Hammer" *MIT Lincoln Laboratory Journal*, 16(2) (To be published).
2. J.J. Garstka, "Network Centric Operations Conceptual Framework: Version 1.0," Office of Force Transformation, Washington, DC (November 2003).
3. W.L. Perry, D. Signori, and J. Boon, *Exploring Information Superiority: A Methodology for Measuring the Quality of Information and its Impact on Shared Awareness*, Rand Corporation, CA, ISBN 0-8330-3489-8 (2003).
4. J. Moffat, "Complexity Theory and Network Centric Warfare," DoD Command and Control Research Program, Washington, DC (2003).
5. S. Kadambe and C. Daniell, "Theoretic Based Performance Analysis of Distributed Sensor Network," HRL Labs, Malibu, CA, Final Technical Report, AFRL-IF-RS-TR-2003-231 (1 October 2003).
6. M.R. Endsley, "Toward a Theory of Situation Awareness in Dynamic Systems," *Human Factors*, 37(1): 32–64 (1995).
7. M.R. Endsley, "Measurement of Situation Awareness in Dynamic Systems," *Human Factors*, (37)1:65–84 (1995).
8. S.M. Mahoney, K.B. Laskey, E.J. Wright, and K.C. Ng, "Measuring Performance for Situation Assessment," *Proc. of 2000 MSS National Symposium on Sensor and Data Fusion*, San Antonio, TX (June 2000).
9. S-A. Knight and J. Burn, "Developing a Framework for Assessing Information Quality on the World Wide Web," *Informing Science Journal*, Volume 8 (2005).
10. E.T. Jaynes, *Probability Theory: The Logic of Science*, ed. G. Larry Bretthorst, Cambridge University Press, Cambridge, UK (April 2003).
11. L. P. Kaelbling, T. Oates, N. Hernandez, and S. Finney, "Learning in Worlds with Objects," *AAAI Stanford Spring Symposium on Learning Grounded Representations*, American Association for Artificial Intelligence (2001).
12. J. Campion, "Decide Now–Ditch Decision Making," Command and Control Research and Technology Symposium, San Diego, CA (June 2004).

REPORT DOCUMENTATION PAGE				Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.					
1. REPORT DATE 31 January 2007		2. REPORT TYPE Technical		3. DATES COVERED (From - To)	
4. TITLE AND SUBTITLE Performance Assessment of Intelligence, Surveillance, and Reconnaissance (ISR) Enterprise Workflows				5a. CONTRACT NUMBER FA8721-05-C-0002	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER 1	
6. AUTHOR(S) M.B. Hurley P.B. Jones				5d. PROJECT NUMBER	
				5e. TASK NUMBER 77411	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) MIT Lincoln Laboratory 244 Wood Street Lexington, MA 02420-9108				8. PERFORMING ORGANIZATION REPORT NUMBER TR-1118	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) Department of the Navy Office of Naval Research One Liberty Center 875 North Randolph Street, Suite 1425 Arlington, VA 22203-1995				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S) ESC-TR-2006-075	
12. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT This report presents an end-to-end assessment framework for C4ISR enterprises that identifies potential FOMs and thereby eliminates a significant fraction of the uncertainty the currently exists on how to assess the performance of these systems. The framework delineates how measurements from laboratory tests, operational experiments, and field deployments can be analyzed to produce accurate, quantitative descriptions of system performance. These FOMs not only inform the development and acquisition processes but also allow military users to ensure that they are optimizing the information enterprise to achieve improved operational capability. In fact, active participation by the military user community in defining appropriate FOMs is essential to providing users with systems that meet their operational needs. An analysis of a simple ISR enterprise through simulation is conducted to show how FOMs might be calculated and used to assess the relative merit of different communications architectures for the construction of a common operational picture (COP) among a collection of distributed sensors.					
15. SUBJECT TERMS ISR enterprise, performance assessment, information theory, probability theory, measures of performance, measures of effectiveness					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT Same as report	18. NUMBER OF PAGES 49	19a. NAME OF RESPONSIBLE PERSON
a. REPORT Unclassified	b. ABSTRACT Unclassified	c. THIS PAGE Unclassified			19b. TELEPHONE NUMBER (include area code)